2017

Talker identification is not improved by lexical access in the absence of familiar phonology

McLaughlin, Deirdre

BOSTON UNIVERSITY

SARGENT COLLEGE OF HEALTH & REHABILITATION SCIENCES

Thesis

TALKER IDENTIFICATION IS NOT IMPROVED BY LEXICAL ACCESS IN THE ABSENCE OF FAMILIAR PHONOLOGY

by

DEIRDRE McLAUGHLIN

B.S., Boston University, 2015

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2017

Approved by

First Reader

Tyler K. Perrachione, Ph.D.
Assistant Professor of Speech, Language, and Hearing Sciences

Second Reader

Sudha Arunachalam, Ph.D.
Assistant Professor of Speech, Language, and Hearing Sciences

Third Reader

Charles Chang, Ph.D.
Assistant Professor of Linguistics

ACKNOWLEDGMENTS

I would like to thank my committee members, Dr. Sudha Arunachalam and Dr. Charles Chang, for their guidance with this project.

I would like to thank the students and faculty members of the speech, language, and hearing sciences and speech-language pathology programs who offered me tremendous support over the last six years and who taught me how to be a speech-language pathologist. A special thank you to my closest friends, Madeline Alexander and Alyssa Mignone, who always support and encourage me.

I would like to thank the members of the CNRLab who helped me on this project. Thank you, Cissy Cheng for helping me generate sentences that were plausible semantically, syntactically, and phonetically in both English and Mandarin. I would never have been able to do this without your expert guidance. Thank you, Sara Dougherty, Lauren Gustainis, Jennifer Golditch, Michelle Lee, Jonathan Mirsky, Andrea Chang, Cassandra Chan, and Emily Thurston for helping me record participants, edit recordings, and run participants. Thank you, Ja Young Choi and Terri Scott for always answering my programming questions and helping me become a better more independent programmer.

Thank you, Liz Pettiti, you have helped me in so many ways throughout my time at BU. You encouraged me from the time I was an undergraduate to pursue a thesis and I think without your support I may never have done it.

An immense thank you to my mentor, Dr. Tyler Perrachione. You taught me to be the best scientist I can be. I couldn't have had a more dedicated, patient, or enthusiastic teacher. I can never repay the opportunities you gave me or the skills you taught me. If I

ever become a clinician-scientist someday, it will be in large part because of your influence.

Finally, I would like to thank my family, especially my parents, without whom I could never have imagined attending Boston University. Thanks for always encouraging me and making everything seem possible.

TALKER IDENTIFICATION IS NOT IMPROVED BY LEXICAL ACCESS IN THE ABSENCE OF FAMILIAR PHONOLOGY

DEIRDRE McLAUGHLIN

ABSTRACT

Listeners identify talkers more accurately when they are familiar with both the sounds and words of the language being spoken. It is unknown whether lexical information alone can facilitate talker identification in the absence of familiar phonology. To dissociate the roles of familiar words and phonology, we developed English-Mandarin "hybrid" sentences, spoken in Mandarin, which can be convincingly coerced to sound like English when presented with corresponding subtitles (e.g., "wei4 gou3 chi1 kao3 li2 zhi1" becomes "we go to college"). Across two experiments, listeners learned to identify talkers in three conditions: listeners' native language (English), an unfamiliar, foreign language (Mandarin), and a foreign language paired with subtitles that primed native language lexical access (subtitled Mandarin). In Experiment 1 listeners underwent a single session of talker identity training; in Experiment 2 listeners completed three days of training. Talkers in a foreign language were identified no better when native language lexical representations were primed (subtitled Mandarin) than from foreign-language speech alone, regardless of whether they had received one or three days of talker identity training. These results suggest that the facilitatory effect of lexical access on talker identification depends on the availability of familiar phonological forms.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Talker identification, or the process of identifying a speaker by the sound of their voice, is an important social and perceptual skill. Researchers have consistently demonstrated that talker identification is functionally integrated with speech perception; however, the sources of information that underlie this integration are at present unknown. In cross-language experiments, listeners perform consistently better at identifying speakers in their native language than in an unfamiliar foreign language, a phenomenon known as the *language-familiarity effect* in talker identification (Thompson, 1987; Goggin, Thompson, Strube, & Simental, 1991; Perrachione & Wong, 2007; reviewed in Perrachione, in press). Likewise, experiments in speech perception have demonstrated that talker variability affects speech processing (Mullenix & Pisoni, 1990, Green, Tomiak, & Kuhl, 1997). The effect of language on processing talker identity demonstrates the bidirectional relationship between linguistic and social perceptual systems: Listeners are able to both resolve talker variability in order to arrive at an underlying linguistic message and to employ an underlying linguistic representation in order to more accurately identify a speaker by the sound of their voice (Kuhl, 2011).

Although the relationship between language familiarity and talker identification ability has been established through a body of scientific work (reviewed in Perrachione, in press), the cognitive model that best explains the interaction between these two sources of information has remained elusive. Some authors have suggested that the language-familiarity effect results from linguistic processing, in which listeners gain access to voice identity-relevant information by processing and representing speech at the level of

linguistic units such as words (Perrachione, Del Tufo, & Gabrieli, 2011; Perrachione, Dougherty, McLaughlin, & Lember, 2015). Other authors have suggested that the language-familiarity effect results only from acoustic-phonetic processing, in which listeners gain access to voice identity-related information by processing speech with respect to the phonetic patterns of their native language (Fleming, Giordano, Caldara, & Belin, 2015; Cutler, 2015). Although both sources of information – acoustic-phonetic and linguistic – have been found to facilitate native-language talker identification, whether these sources of information contribute independently to this ability, or whether there is a bi-directional or hierarchical dependence between these representations – has not been explored.

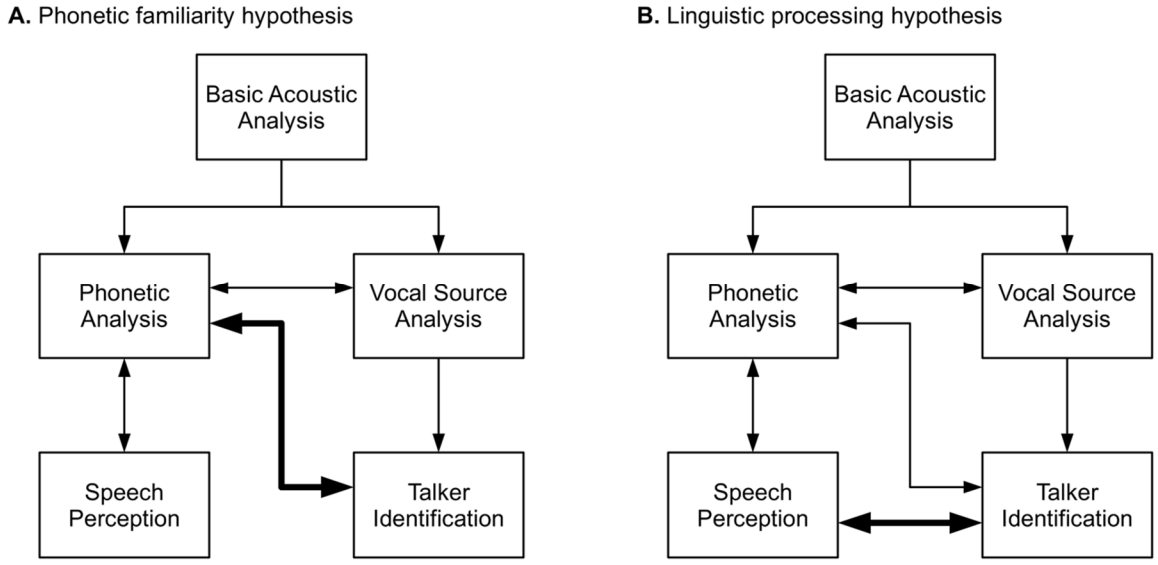

Figure 1: Hypothesized Models of Talker Identification (Perrachione (in press)): This figure depicts the two current hypothesized models of talker identification as reviewed in Perrachione (in press): a) the phonetic-familiarity hypothesis and b) the linguistic processing hypothesis. The phonetic familiarity model proposes that talker identification can be performed at low levels of phonetic processing and can be independent of speech processing. The linguistic processing model proposes that talker identification is facilitated by speech perception (as reviewed in Perrachione (in press)).

There is evidence to suggest that the language-familiarity effect in talker identification can be facilitated by processing familiar acoustic-phonetic features in the absence of lexical processing. First, listeners subjectively judged talkers as sounding more dissimilar in their native language than in an unfamiliar foreign language when the speech signal was time-reversed and therefore incomprehensible (Fleming, Giordano, Caldara, & Belin, 2015). This finding supports a model in which the language-familiarity effect arises from processing the low-level speech phonetics and acoustics that are preserved in time-reversed speech. Second, self-reported monolingual English listeners from Canada were better at talker identification in French than were monolingual English listeners from the United States. These findings suggest that, because listeners from Montreal have been passively exposed to the sound structure of an unfamiliar language (but allegedly not higher-level linguistic structure like lexical items), this passive exposure to sound structure may be sufficient to reduce the language-familiarity effect (Orena, Theodore, & Polka, 2015). Third, the language familiarity effect has been demonstrated in infants as young as 7 months, which is arguably before the establishment of higher level linguistic representations like words (Johnson, Westrek, Nazzi, & Cutler, 2011). Finally, a greater language familiarity effect is found between languages that are phonologically dissimilar than languages that are phonologically similar, which suggests that the effect is related to phonological processing of phonetic differences (Zarate, Tian, Woods, & Poeppel, 2015). Collectively, these results indicate evidence that phonological processing at the level of phonetics and acoustics is important for talker identification performance.

There is also evidence to suggest that the language-familiarity effect in talker identification is facilitated by higher-level linguistic processing – in particular, lexical access. Several studies have shown that talker identification abilities improve as a function of the linguistic information available. Listeners performed more accurately at identifying talkers as the amount of phonological information increased from a vowel to a word to a sentence (Pollack, Pickett, & Sumby, 1954; Bricker & Pruzansky, 1966; Zarate, Tian, Woods, & Poeppel, 2015). Talker identification is also improved as the quantity of known as opposed to novel words increases: Listeners perform significantly better at identifying talkers from speech consisting of known words compared to nonsense speech with a high-probability native-language phonological structure (Perrachione, Dougherty, McLaughlin, & Lember, 2015; Goggin et al., 1991; Zarate, Tian, Woods, & Poeppel, 2015; Xie & Myers, 2015). Listeners also identify talkers more accurately in their native language compared to a foreign language when the lexical content of the speech is repeated, revealing that consistent (but unknown) speech content confers no talker identification benefit in a foreign language (McLaughlin, Dougherty, Lember, & Perrachione, 2015).

Contrary to the findings when listeners subjectively judge talker similarity (Fleming, Giordano, Caldara, & Belin, 2015), listeners do not demonstrate a language familiarity effect when identifying voices from time-reversed speech (Perrachione, McLaughlin, Dougherty, & Lember, 2015; Dougherty, McLaughlin, & Perrachione, 2015). This discrepancy suggests that phonological and lexical processing may play a greater role in active talker identification, and therefore make a greater contribution to the magnitude of

the language-familiarity effect, than they do for tasks like judging talker similarity. This discrepancy is further paralleled in the effect size of the language familiarity effect in these two types of tasks: When listeners rate talker similarity from time-reversed speech, language familiarity results in a difference of 3% in similarity judgments, whereas when listeners learn to identify voices, language-familiarity confers a benefit of up to 24% (Fleming, Giordano, Caldara, & Belin, 2015; reviewed in Perrachione, in press).

In the current study, we explored the contribution of lexical access vs. acoustic-phonetic processing further by examining whether access to familiar words in a foreign language condition (where familiar sounds are not present) would facilitate talker identification performance. In two experiments, listeners heard spoken sentences in Mandarin that could be coerced to sound like English when presented with subtitles that prime lexical expectations in speech processing. These sentences were carefully designed to create semantically and syntactically plausible sentences in both languages with the presence of subtitle lexical priming. This coercion was plausible given the pop culture phenomenon of *mondegreens* –, or perceiving foreign language speech or songs to be native language words when presented with native language subtitles, and the current literature on lexical biases in speech perception (Liberman, 2007).

Speech perception research has demonstrated numerous circumstances in which top-down expectations about words can influence listeners' speech processing. First, expectation-biased perception has been observed in categorical perception of ambiguous phonemes. When listeners are presented with two words in which their initial sounds varied on a phonemic continuum, they are more likely to disregard competing acoustic

information in favor of perceiving a real word (Ganong, 1980). This lexical bias shifts the phonetic continuum in favor of the sound that results in a real word. Expectation-based perceptual biases also extend to richer phonetic contexts such as sentences. Listeners' perception of vocoded sentences, where spectral information is removed from the speech signal, is better when they expected key content words from the sentence (Davis, Johnsrude, Hervis-Adelman, Taylor & McGettigan, 2005). Lexical perception is also enhanced in the presence of subtitles that match the phonetic context. Listeners more accurately recognized vocoded speech in a test phase with matching as opposed to mismatching subtitles during training phases (Sohoglu & Davis, 2016). In addition, when listeners are presented with videos in a second language and are provided subtitles, native-language subtitles appear to create lexical interference but foreign-language subtitles assist speech learning by indicating which words and sounds are being spoken (Mitterer & McQueen, 2009).

In the present study, we tested the hypothesis that providing lexical primes via subtitles in a foreign-language condition would improve talker identification accuracy compared to a condition in which no primes were presented. If this were the case, lexical access to familiar words facilitates talker identification independent of access to familiar sounds, indicating that linguistic processing facilitates talker identification. Alternatively, lexical primes may have no effect on talker identification abilities; this would suggest that processing on lower levels contributes to talker identification, indicating that this process is hierarchical. Across two experiments involving different amounts of training, we found that lexical priming does not appear to improve talker identification in the

absence of familiar phonological information.

EXPERIMENT 1: PRIMING LEXICAL REPRESENTATIONS DURING FOREIGN LANGUAGE TALKER IDENTIFICATION

2.1 Methods

2.1.1 Participants

Native speakers of American-English completed this study (N = 16, age 18–28 years, M = 20.75, 13 female). Inclusion criteria required participants to have a self-reported history free from speech, language, or hearing problems and no prior experience with Mandarin. This study was approved and overseen by the Institutional Review Board at Boston University. Participants provided written informed consent and received monetary compensation for their participation.

2.1.2 Stimuli

Twenty *English-Mandarin hybrid sentences* were designed for this experiment. Each sentence was required to meet the following criteria: (1) be composed of Mandarin words with sounds that could be coerced to be perceived as English and (2) be syntactically and semantically plausible in both language conditions. Words and short phrases in each sentence were generated or selected from existing corpora (phonetically balanced sentence sets by Fu, Zu, & Wang, 2011) using knowledge of the phonotactic properties of English, Mandarin, and Mandarin-accented English, as well as the patterns of perception of Mandarin phonemes by English speakers (Tsao, Liu, & Kuhl, 2006).

The English-Mandarin hybrid sentences were recorded by ten female native speakers of Mandarin (age 19–27, M = 23), and their corresponding English sentences

were recorded by ten female native speakers of American English (age 19–29, M = 22.3). Both groups of talkers were without distinctive regional accents. Recordings were made in quiet in a sound attenuated booth using a Shure MX153 earset microphone, a Behringer Ultragain Pro MIC2200 2-channel tube microphone preamplifier, and Roland Quad Capture USB audio interface with a sampling rate of 44.1 kHz and 16-bit digitization in Praat RMS amplitude. Each sentence was RMS-amplitude normalized to 65 dB SPL using *Praat version 5.3.63* (http://praat.org).

Mandarin	**English**
陪你晚到了 pʰeɪ ni wɑn tɑʊ lə péi nǐ wǎn dào le *With you, I was late.*	Pay me one dollar. pʰeɪ mi wʌn dɑlɚ
喂狗吃烤荔枝 weɪ koʊ tʂʰɨ kʰɑʊ li tʂɨ wèi gǒu chī kǎo lì zhī *Feed the dog grilled lychees.*	We go to college. wi goʊ tʰu kɑlədʒ
妈妈喜欢芒果 mɑ mɑ ɕi xwɑn mɑʊ tsɨ mā mā xǐ huān mào zi *Mother likes the hat.*	Mama sees one mouse. mɑmə siz wʌn mɑʊs.

Table 1: Experimental Stimuli This table contains examples of the English Mandarin hybrid sentences that were designed for this project. The first column contains the Mandarin version of the sentences with Mandarin characters, phonetic transcription, pinyin, and English translation. The second column contains the English sentences, with English orthography and phonetic transcription.

Because some voices are inherently more distinctive than others, stimuli were extensively piloted prior to running this experiment in order to develop within-language voice sets that were equally identifiable. Additional listeners learned to identify different

groupings of these voices, allowing us to balance listeners' within-language accuracy between the two sets of talkers. These pilot tasks ensured that, absent the lexical priming manipulation in the actual experiment, mean accuracy did not differ for each set of talkers. Furthermore, the two sets of talkers in each language were also counterbalanced across experiment conditions.

2.1.3 Procedure

The experiment consisted of a 2×2 factorial design in which the language being spoken (English or Mandarin) and the presence of top-down lexical priming (with or without subtitles) were varied. This resulted in four conditions: (1) English with subtitles, (2) English without subtitles, (3) Mandarin with subtitles, (4) Mandarin without subtitles. Participants completed all conditions of the experiments in a single session. The order of conditions and the within-language talker groupings were counterbalanced across participants. The sentences and talkers were not repeated within or between experimental conditions for each participant. In conditions with subtitles, the priming text appeared two seconds before the presentation of the recorded sentence in order for listeners to have the opportunity to activate the target lexical representations before hearing the sentence. In the conditions without subtitles, a blank screen appeared for two seconds at the beginning of each trial.

Each condition of the experiment consisted of three phases: (1) a familiarization phase, (2) an active practice phase, and (3) a test phase. This design has been used in previous studies to train and test talker identification in a single experimental session (Perrachione & Wong, 2007; Perrachione, Del Tufo, & Gabrieli, 2011; Perrachione,

Dougherty, McLaughlin, & Lember, 2015). The experiment was programmed using PsychoPy (Peirce, 2007). Participants listened to stimuli using Sennheiser HD 380 Pro headphones and selected the talker using a keypad.

In the training phase of the experiment, participants received passive exposure and active practice at identifying the talkers in each group. In the passive exposure phase, participants listened to each voice in isolation while the corresponding avatar and number appeared on the screen. Participants were not prompted to make a response during this phase. In this phase, participants heard each speaker say five sentences twice for a total of 50 trials (5 talkers x 5 sentences x 2 repetitions). In the active practice phase, participants were prompted to match a talker's voice with a cartoon avatar and talker number from a field of five talkers. Participants were provided corrective feedback in this phase that indicated whether their choice was correct or incorrect and who the correct speaker was. In this phase, participants practiced identifying each speaker by the sound of their voice when they were saying five sentences for a total of 50 trials (5 talkers x 5 sentences x 2 repetitions). The trials in the training phase were blocked by sentence. Therefore, participants listened to each talker say the same sentence and then immediately after practiced choosing each talker from a field of five.

In the test phase, participants were prompted to match the talker's voice with its corresponding cartoon avatar and talker number without feedback. In this phase, participants were prompted to identify each talker saying each sentence twice for a total of 50 trials (5 talkers x 5 sentences x 2 repetitions). The presentation of talkers and sentences during the test was randomized.

2.1.4 Data Analysis

Participants' mean accuracy (the percentage of correct responses of the total number of trials) was calculated for each condition. Participants' accuracy in each condition was analyzed using *R* version 3.3.2 (https://cran.r-project.org/) using a repeated measures ANOVA implemented in the "ez" library. Within-group factors included language condition and priming. A paired t-test was also calculated to analyze the potential effect of priming within language conditions (English with and without priming; Mandarin with and without priming).

2.2 Results

All listeners demonstrated a language-familiarity effect such that they performed with significantly greater accuracy when identifying talkers in the native language conditions (M ± SE, English: 78.0% ± 16.8%; English with priming: M=78.0% ± 15.5%) than in unfamiliar foreign language conditions (Mandarin: 38.6% ± 17.2%; Mandarin with Priming: 42.5% ± 17.9%) (main effect of language; $F_{1,15} = 250.66$, $p < 9.04 \times 10^{-11}$, $\eta^2_G = 0.57$). Listeners' performance in both native and foreign language conditions did not significantly differ when there was lexical priming available (English, Mandarin with priming) than when lexical priming was not available (no main effect of priming; $F_{1,15} = 0.32$, $p = 0.58$, $\eta^2_G = 0.0035$). Listeners did not perform significantly better in a foreign language condition when lexical priming was available than when lexical priming was absent (paired $t_{15} = -0.90$, $p = 0.38$). There was also no interaction effect between language and lexical priming, suggesting priming did not differentially facilitate talker identification in one language vs. the other ($F_{1,15} = 0.81$, $p = 0.382$, $\eta^2_G = 0.0035$).

Together, these results suggest that expectations about the lexical content of speech do not improve talker identification when listening to either a native or, importantly, a foreign language.

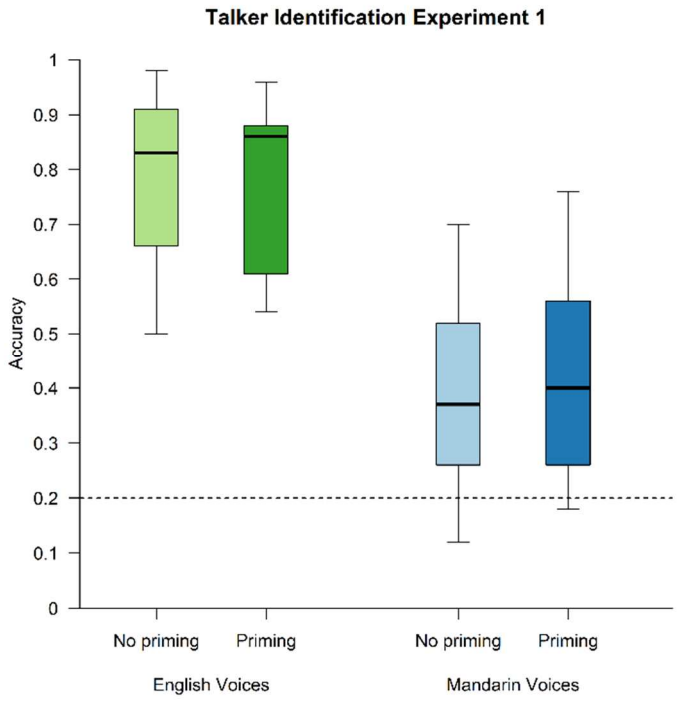

Figure 2: The boxplot shows the participants' mean accuracy by condition. The dashed line represents chance-level accuracy (20%).

2.3 Discussion

As in previous experiments, listeners performed better at identifying talkers in a native language than in a foreign language. However, listeners did not perform better in a foreign language when provided with expectations about hearing familiar words. This suggests that listeners are not able to gain additional information about talker identity from familiar words in the presence of an unfamiliar phonology.

There were several topics of investigation not explored in this experiment that we wished to explore further. First, it was possible that in Experiment 1 listeners did not

receive adequate exposure to lexical primes and required additional time or training to utilize lexical forms in a foreign language condition. In a previous experiment with bilingual participants, listeners exhibited the language-familiarity effect in favor of their native (vs. second) language on the first day of training, but the magnitude of the language difference was attenuated and eventually nullified after additional days of training (Perrachione & Wong, 2007). In this way, it may be the case that further training on foreign-language voices in the presence of lexical primes may be necessary to take advantage of this additional information source and improve talker identification performance. Additionally, participants were tested only on trained sentences. This is different than other previous experimental paradigms where untrained sentences were introduced into the test phase (5 trained, 5 untrained) (Perrachione, Dougherty, McLaughlin, & Lember, 2015). It may also be the case that the information sources made available during lexical priming will be more beneficial in facilitating talker identification from untrained sentences – a condition in which accuracy typically decreases (Perrachione & Wong, 2007; McLaughlin, Dougherty, Lember, & Perrachione, 2015).

Given these considerations, we repeated the lexical priming manipulation in Experiment 2, in which participants were exposed to language conditions with and without lexical primes for repeated training sessions across three consecutive days. In addition, to test the role of lexical access in participants' ability to generalize to untrained exemplars, both trained and untrained sentences were included in the test phase of the experiment for each training day.

EXPERIMENT 2: TRAINING FOREIGN-LANGUAGE TALKER IDENTIFICATION WITH LEXICAL PRIMING

3.1 Methods

3.1.1 Participants

Native speakers of American-English completed this study (N = 18, age 18–27 years, M = 20.5, 14 female). Inclusion criteria were the same as Experiment 1. Inclusion criteria for Experiment 2 also required that participants perform with greater than chance accuracy in any condition. Four additional participants completed the study but were excluded due to failure to meet the accuracy criterion. This study was approved and overseen by the Institutional Review Board at Boston University. Participants provided written informed consent and received monetary compensation for their participation. Participants in Experiment 2 did not participate in Experiment 1.

3.1.2 Stimuli

Participants were presented with sentence stimuli from three different corpora in three different conditions: English, Mandarin with subtitles to prime a target English gloss, and Mandarin without subtitles. In the English condition, listeners heard talkers saying phonetically balanced sentences in English, drawn from a previous talker identification study (McLaughlin, Dougherty, Lember, & Perrachione, 2015). In the Mandarin conditions, listeners heard sentences drawn from a set of phonetically balanced Mandarin sentences (Fu, Zhu, & Wang, 2011). In the Mandarin condition with subtitles to prime an intended English gloss, listeners heard talkers saying the English-Mandarin hybrid sentences that were designed for Experiment 1. Recordings from a group of five

native speakers of American English (age 20–29, M=23.4) and a group of ten native Mandarin speakers (age 19–27, M = 23) were selected from existing corpora of recordings. These groupings were piloted in previous experiments to ensure that no talker is inherently more identifiable in a group than another talker (McLaughlin, Dougherty, Lember, & Perrachione, 2015).

3.1.3 Procedure

Participants learned to identify talkers' voices in three training and testing sessions on consecutive days. Participants learned a different group of voices in each of the three conditions: English, Mandarin with subtitle primes, and Mandarin. In the English and Mandarin conditions, there were no subtitles. As in the previous experiment, the priming text appeared two seconds before the presentation of the recorded sentence in order for the listener to have the opportunity to read the sentence and activate the target lexical representations before hearing the corresponding speech. In the conditions without subtitles (English, Mandarin), a blank screen appeared for two seconds at the beginning of each trial, such that the timing of each condition was the same.

Participants completed all conditions of the experiments in every session. The order of conditions was counterbalanced across participants, but kept the same within participant across days. Voice groupings were counterbalanced across participants in the two Mandarin conditions.

The conditions contained the same experimental phases as Experiment 1: (1) a familiarization phase, (2) an active practice phase, and (3) a test phase. Across sessions, participants trained on the same five sentences during the familiarization and active

practice stages. During the test phase of each session, participants were asked to identify voices speaking both the sentences that they had been trained on and five new sentences. The new sentences were included to assess how well the participant's knowledge of that speaker's voice generalized to untrained sentences.

3.1.4 Data Analysis

As in Experiment 1, participants' mean accuracy or the average percentage of correct responses of the total number of trials was calculated for each condition in experiment 2. Participants' accuracy in each condition was analyzed using *R* version 3.3.2 (https://cran.r-project.org/) using the repeated measures ANOVA implemented in the "ez" library.

3.2 Results

A repeated measures ANOVA comparing participants' performance on all three conditions revealed a significant effect of condition ($F_{2,34} = 106.37$, $p < 2.33 \times 10^{-15}$, $\eta^2_G = 0.82$) and training day ($F_{1,17} = 47.11$, $p < 2.75 \times 10^{-6}$, $\eta^2_G = 0.16$) such that participants' performance in English was significantly greater than in Mandarin, and that performance in all conditions improved across training days. There was not an interaction effect between condition and training day ($F_{2,34} = 1.70$, $p = 0.20$, $\eta^2_G = 0.02$) such that the rate of learning did not appear to differ across the three conditions overall

A second repeated measures ANOVA comparing participants' performance on only the Mandarin conditions was performed, revealing a significant effect of training day ($F_{1,17} = 19.29$, $p < 0.0005$, $\eta^2_G = 0.12$) such that participant's performance improved significantly for foreign-language voices across the three sessions. There was not a

significant effect for condition ($F_{1,17} = 0.95$, $p = 0.34$, $\eta^2_G = 0.04$), such that listeners did no better in Mandarin conditions with lexical primes than in Mandarin conditions without lexical primes. There was also not an interaction effect between training day and condition ($F_{1,17} = 1.38$, $p = 0.26$, $\eta^2_G = 0.01$), such that the rate of talker learning was not greater in the Mandarin condition with lexical primes than the Mandarin condition without the lexical primes.

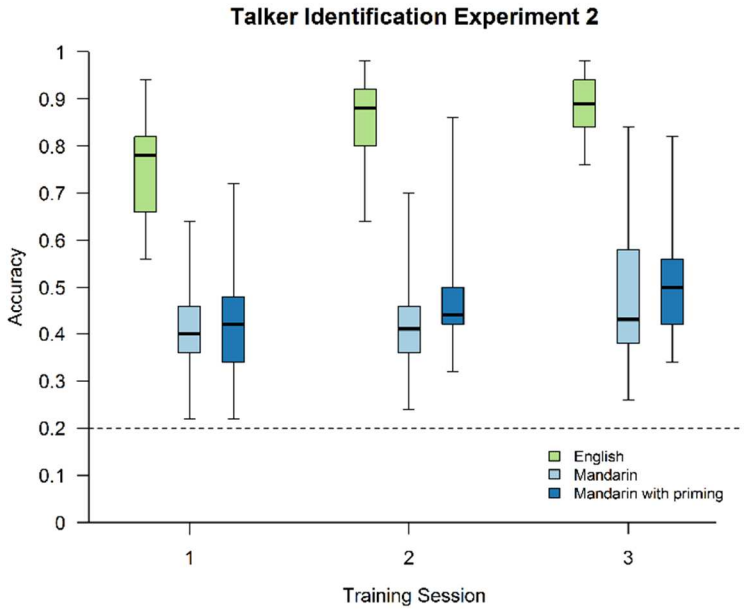

Figure 3: The boxplot shows the participants' mean accuracy by condition and training day. The dashed line represents chance-level accuracy (20%).

A third repeated measures ANOVA comparing participants' performance on trained versus untrained sentences across training days in the Mandarin conditions revealed a significant effect of training day ($F_{1,17} = 19.29$, $p < 0.0004$, $\eta^2_G = 0.081$) and of sentence familiarity ($F_{1,17} = 18.47$, $p < 0.0005$, $\eta^2_G = 0.09$) such that participants' accuracy on trained and untrained sentences improved over time and that participants performed

better on trained sentences relative to untrained sentences overall. As before, there was not a significant effect of condition ($F_{1,17} = 0.95$, $p = 0.34$, $\eta^2_G = 0.03$). There were also no significant interaction effects (condition × training day × sentence exposure: $F_{1,17} = 0.06$, $p = 0.81$, $\eta^2_G = 0.00039$; condition × training day: $F_{1,17} = 1.38$, $p = 0.26$, $\eta^2_G = 0.0089$; condition × sentence exposure: $F_{1,17} = 1.08$, $p = 0.31$, $\eta^2_G = 0.0022$; and training day × sentence exposure: $F_{1,17} = 1.35$, $p = 0.26$, $\eta^2_G = 0.0066$), indicating that the effect of sentence familiarity did not differ across conditions or training days, and that the rate of improvement did not differ based on any combination of variables.

3.3 Discussion

As in Experiment 1, participants did not perform significantly better in a foreign language condition when primed with known lexical representations than in a foreign language condition with no known lexical representations, even when provided with multiple exposures to prime lexical expectations across several sequential days of training. Participants reliably demonstrated the language familiarity effect which is consistent with all previous findings (e.g., Thompson, 1987; Goggin, Thompson, Strube, & Simental, 1991; Perrachione & Wong, 2007). Participants also improved their talker identification performance in both language conditions across training days but did not reduce the magnitude of the language-familiarity effect with additional training, which is also consistent with previous studies reporting no training-based reduction in the language-familiarity effect for monolingual listeners (Perrachione & Wong, 2007). Finally, listeners performed better on trained sentences than on untrained sentences in a native language, which is consistent with the benefit of lexical repetition in native

language conditions viewed in other studies (McLaughlin, Dougherty, Lember, & Perrachione, 2015). The results again demonstrate that participants did not use known lexical representations in the absence of a familiar phonology to identify voices in a second language in Experiment 2, even when provided multiple training sessions from which to learn to take advantage of this potential information source.

GENERAL DISCUSSION

In two experiments, we found that the magnitude of the language familiarity effect was not reduced even when listeners had access to familiar words in a foreign language condition via priming for native language lexical expectations. In Experiment 1, listeners' performance did not improve as a result of primed lexical representations in either the native or foreign language conditions during a single training session. Likewise in Experiment 2, even though listeners were given multiple training sessions to learn to draw upon lexical primes as a way to improve their talker identification performance, we observed essentially the same pattern of results as in Experiment 1. Taken together, these results suggest that the language-familiarity effect in talker identification is not driven by access to familiar words in the absence of familiar sounds.

These results help refine our models for the cognitive and perceptual processes underlying talker identification. Currently, there is evidence to suggest that speech processing and talker identification interact, but it is unknown at what level of linguistic processing this interaction occurs. Some research has indicated the importance of acoustic-phonetic processing as a basis for this effect (Johnson, Westrek, Nazzi, & Cutler, 2011; Fleming, Giordano, Caldara, & Belin, 2015; Orena, Theodore, & Polka, 2015; Zarate, Tian, Woods, & Poeppel, 2015). On the other hand, others have argued that lexical processing also plays a role (Perrachione & Wong, 2007; Perrachione, Del Tufo, & Gabrieli, 2011; Perrachione, Dougherty, McLaughlin, & Lember, 2015; McLaughlin, Dougherty, Lember, & Perrachione, 2015). The present results reveal additional nuance to the role of lexical processing in a more complete model of talker identification – that

is, lexical processing only appears to play a facilitatory role in the presence of familiar acoustic-phonetic information. When familiar phonetic features are unavailable, it does not appear that listeners are able to make use of lexical access to facilitate talker identification.

Ultimately, these results suggest that the cognitive processes involved in talker identification are supported by a hierarchy of perceptual cues, each of which is likely to depend on successful processing of the previous level. At the lowest level, listeners extract prelinguistic and relatively invariant information about a talker's voice such as fundamental frequency (f_0) and f_0 range, formant dispersion and vocal tract length, and voice quality. Beyond global acoustic properties, listeners gain additional information from acoustic-phonetic features when such features are familiar due to long-term linguistic experience. Naturally, access to phonetic information depends on successful low-level processing and encoding of the auditory signal. Finally, listeners gain additional information about a talker's identity from processing higher-level linguistic information such as through lexical access and memories for words. However, the present experiments suggest that access to this level of information depends on successfully parsing and representing the prior (acoustic-phonetic) level. In all the previous talker identification experiments that have demonstrated beneficial effects of lexical access, lexical information was manipulated in the presence of familiar acoustic-phonetic and phonological structures (Pollack, Pickett, & Sumby, 1954; Bricker & Pruzansky, 1966; Zarate, Tian, Woods, & Poeppel, 2015; Perrachione, Dougherty, McLaughlin, & Lember, 2015, Xie & Myers, 2015a, McLaughlin, Dougherty, Lember, & Perrachione, 2015).

Although these experiments showed in various ways that access to word-level representations can improve listeners' abilities to identify voices, they did not explore whether such facilitation depended on successful processing of a lower-level of information, namely the presence of familiar phonology.

The present results provide new insight into literature on influences of unfamiliar regional and social accents on talker identification, particularly accented talker identification in listeners' native language. Listeners have consistently been shown to perform worse at identifying talkers speaking an in unfamiliar social or regional accent in native language conditions (Thompson, 1987; Goggin, Thompson, Strube, & Simental, 1991; Doty, 1998; Kerstholt, Jansen, Amelsvoort, & Broeders, 2006; Perrachione, Chiao, & Wong, 2010; Stevenage, Clarke, & MacNeill, 2012). In all these cases, listeners putatively had access to lexical information to some extent, since the linguistic content was familiar. However, while talker identification is poorer in an unfamiliar accent than in a familiar one – likely due to less experience with the characteristic distributions of the phonetic features in the unfamiliar accents – across studies, performance in an unfamiliar accent of a native language remains much better than in a foreign language, where both the linguistic and phonetic features are unfamiliar. In this way, it was unclear whether priming access to familiar words (in the absence of familiar phonology) would nonetheless improve talker identification over a fully foreign language condition, even if listeners' performance still did not reach the level of the native language condition (since voices speaking accented L1 speech are still much better identified than L2 voices). A principal contribution of the present experiment is to show that there is indeed a

dependency relationship between familiar words and familiar sounds – the former is only beneficial in the presence of the latter, particularly when the latter is very unfamiliar.

Although straightforward interpretation of these results indicates that talker identification is not improved by spoken language recognition in the absence of a familiar phonology, there are possible limitations of these findings that should be considered. First, it is possible that listeners in the Mandarin-with-priming condition did not always convincingly perceive the sentences as English. Although this is a possibility, through piloting and across experimental sessions, listeners qualitatively reported that the sentences with subtitles did sound like foreign-accented speech. It may be possible to quantify listeners' perception of the sentences as English when heard in the presence of lexical priming through additional experiments, such as a sentence transcription task following the Mandarin with priming condition. Correspondingly, there may have been item-specific variability within the set of sentences, such that that some sentences primed English lexical representations more effectively than others. Similarly, it is possible that priming with native language subtitles prior to the presentation of the Mandarin sentence may in some cases have actually negatively impacts listeners' performance because of cognitive demands of reorienting after experiencing a mismatch between their expectation of the sentence and the actual recording. However, given the exceedingly similar means and distributions of the primed and unprimed conditions, we believe such subtle item effects to be unlikely. Additional future analyses of these data will be needed to identify if there are any item-specific effects, such as whether the most convincing sentence primes actually conferred more of a benefit than the less-convincing ones.

Finally, this experiment did not investigate the potential effect of cognitive resources on the talker identification in a second language condition. It is possible that less accurate performance in a foreign language could result from limitations in cognitive resources such as having fewer attentional resources to devote to processing foreign language speech rather than from lack of experience in linguistic processing of that language. Further research is necessary to adjudicate between the potential effect of cognitive resources on talker identification in a familiar or unfamiliar language. Differences in cognitive load notwithstanding, the present results provide important new insight into how various information sources contribute to talker identification because listeners consistently do not benefit from familiar words in the absence of familiar sounds, even when provided multiple exposures of lexical primes across several days of training.

In sum, these experiments suggest that a more complete model of the cognitive processes involved in talker identification includes both acoustic-phonetic and higher-level linguistic processing, and that there is a hierarchical relationship among these linguistic levels, where the facilitatory effects of lexical access in talker identification depends specifically on the availability of familiar acoustic-phonetic features.

APPENDIX

Experiment 1: Mandarin sentences and corresponding English glosses

1. **肚子偶**尔有些痛
   Do it or you sit home
2. **院子**门口被遮住了
   You and the men go pay your jeweler
3. **麻**烦你走这里
   My friend needs some jelly
4. **我**们看到了小平
   Women can do the shopping
5. **紫色礼服干些了**吗
   The silly fool can say llama
6. **我**爱吃大白兔糖
   Why should I buy two tons
7. 妈妈喜欢芒果
   Mama sees one mouse
8. **我**们喜爱松鼠
   They see one baker
9. **他网上**买书
   What did I try to show Joe
10. **我喂美**丽的兔子糖
    What way may lead to the town
11. **幼儿老**师不会特胖
    Your lost sheep weighs two pounds
12. **喂狗吃烤荔枝**
    We go to college
13. **温度有些高**
    When do you see a cow
14. 爱买白松鼠
    Why I might buy some shoes
15. **有你的**饭袋
    You need a fun day
16. **陪你晚到了**
    Pay me one dollar
17. **有你陪着我**们
    You need to pay the woman
18. **我的有肉**丝
    Water your roses
19. **好的构**图技巧
    How to go to the town
20. 护士的新的帽子
    Who should have seen the mouse

Experiment 2: English sentences

1. Granola is best in yogurt.
2. Tots adorn Easter eggs.
3. Telescopes view constellations.
4. Babies laugh happily.
5. Policemen chase criminals.
6. Maps show country boundaries.
7. Puppies bark at passing trucks.
8. Polka dots decorate fabric.
9. Annoying birds chirp noisily.
10. Perennials bloom all year.
11. Textbooks burden backpacks.
12. Calculators solve problems.
13. Broken headphones produce static.
14. Handy rulers draw straight lines.
15. Locks protect valuables.
16. Ice relieves joint pain or swelling.
17. Unwelcome weeds invade lawns.
18. She crafted clever shortcuts.
19. Loud alarms wake roommates.
20. Kangaroos jump along.
21. Shower pipes spray water.
22. Wind pushes against heavy doors.
23. Most flowers grow slowly.
24. Studying improves exam scores.
25. Special coffee mugs are great gifts.
26. Thumb tacks support posters.
27. Bosses manage employees.
28. Plugs supply electricity.
29. City sidewalks dirty shoes.
30. Insulation stops heat loss.

REFERENCES

Bricker, P. D., & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *The Journal of the Acoustical Society of America*, *40*(6), 1441–1449.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222.

Doty, N. D. (1998). The influence of nationality on the accuracy of face and voice recognition. *The American Journal of Psychology*, *111*(2), 191.

Dougherty, S.C., McLaughlin, D.E., & Perrachione, T.K. (2015) "A language familiarity effect for talker identification in forward but not time-reversed speech." *169th Meeting of the Acoustical Society of America* (Pittsburgh, May 2015).

Fleming, D., Giordano, B. L., Caldara, R., & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, *11*(38), 13795–13798.

Fu, Q. J., Zhu, M., & Wang, X. (2011). Development and validation of the Mandarin speech perception test. *The Journal of the Acoustical Society of America,* 129, EL267–EL273.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110.

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, *19*(5), 448–458.

Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics*, *59*(5), 675–692.

IEEE. (1969). IEEE recommended practices for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, 17*, 225–246.

Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, *14*(5), 1002–1011.

Kadam, M. A., Orena, A. J., Theodore, R. M., & Polka, L. (2016). Reading ability influences native and non-native voice recognition, even for unimpaired readers. *The Journal of the Acoustical Society of America*, *139*(1), EL6–EL12.

Kerstholt, J. H., Jansen, N. J., Van Amelsvoort, A. G., & Broeders, A. P. A. (2006). Earwitnesses: Effects of accent, retention and telephone. *Applied Cognitive Psychology*, *20*(2), 187–197.

Kuhl, P. K. (2011). Who's talking? *Science*, *333*(6042), 529–530.

Liberman, M. (2007) Autour-Du-Mondegreens: Bunkum Unbound. *Language Log*. Retrieved from http://itre.cis.upenn.edu/~myl/languagelog/archives/005100.html

McLaughlin, D. E., Dougherty, S. C., Lember, R. A., & Perrachione, T. K. Episodic memory for words enhances the language familiarity effect in talker identification. *18th International Congress of Phonetic Sciences* (Glasgow, August 2015)

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS One*, *4*(11), e7785.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, *47*(4), 379–390.

Orena, A. J., Theodore, R. M., & Polka, L. (2015). Language exposure facilitates talker learning prior to language comprehension, even in adults. *Cognition*, *143*, 36–40.

Perrachione, T.K. (submitted) "Speaker recognition across languages" in S. Frühholz & P. Belin (Eds.), *The Oxford Handbook of Voice Perception*, Oxford: Oxford University Press.

Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. (2011). Human voice recognition depends on language ability. *Science*, *333*(6042), 595.

Perrachione, T. K., Dougherty, S. C., McLaughlin, D. E., & Lember, R. A. The effects of speech perception and speech comprehension on talker identification. *18th International Congress of Phonetic Sciences* (Glasgow, August 2015)

Perrachione, T. K., Chiao, J. Y., & Wong, P. C. (2010). Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition*, *114*(1), 42–55.

Perrachione, T. K., Pierrehumbert, J. B., & Wong, P. (2009). Differential neural contributions to native-and foreign-language talker identification. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1950.

Perrachione, T. K., & Wong, P. C. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia*, *45*(8), 1899–1910.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13.

Pollack, I., Pickett, J. M., & Sumby, W. H. (1954). On the identification of speakers by voice. *The Journal of the Acoustical Society of America*, *26*(3), 403–406.

Stevenage, S. V., Clarke, G., & McNeill, A. (2012). The "other-accent" effect in voice recognition. *Journal of Cognitive Psychology*, *24*(6), 647–653.

Thompson, C. P. (1987). A language effect in voice identification. *Applied Cognitive Psychology, 1*(2), 121–131.

Xie, X., & Fowler, C. A. (2013). Listening with a foreign-accent: The interlanguage speech intelligibility benefit in Mandarin speakers of English. *Journal of Phonetics*, *41*(5), 369–378.

Xie, X., & Myers, E.B. (2015). General language ability predicts talker identification. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio, P. P. (Eds.) (2015). *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Zarate, J. M., Tian, X., Woods, K. J., & Poeppel, D. (2015). Multiple levels of linguistic and paralinguistic features contribute to voice recognition. *Scientific Reports*, *5*.

VITA

[redacted]